Complément au théorème de Simmons sur les probabilités.

Par CHARLES JORDAN à Budapest.

Un événement se produit avec la probabilité p. En n épreuves on s'attend à le voir arriver à peu près np fois. Si np est entier et si $p < \frac{1}{2}$, le théorème de SIMMONS affirme[1]) qu'il y a plus de chances pour l'événement d'arriver moins de np fois que plus de np fois.

Nous nous sommes proposés d'examiner le cas où np n'est pas entier $\left(\text{en maintenant l'hypothèse } p < \frac{1}{2}\right)$ et nous sommes arrivés au résultat partiel suivant:

1. Si np diffère assez peu de l'entier immédiatement inférieur $[np]$, de façon que la différence n'excède pas $\frac{1}{2}$, l'affirmation contenue dans le théorème de SIMMONS subsiste.

2. Si la différence $np-[np]$ égale ou dépasse $1-p$, c'est l'inverse qui a lieu, savoir: l'événement a moins de chances d'arriver moins de np fois que plus de np fois.

Lorsque la différence $np-[np]$ est comprise entre $\frac{1}{2}$ et $1-p$, nous n'avons pas pu élucider complètement le problème; il paraît cependant que le renversement des chances se produit à peu près vers le milieu, c'est-à-dire approximativement quand

$$np-[np]=\frac{1}{2}\left(q+\frac{1}{2}\right).$$

1) T. L. SIMMONS, A New Theorem of Probability, *Proceedings of the London Mathematical Society*, **26** (1894—95), pp. 290—334.

RAGNAR FRISH, Solution d'un Problème de Probabilités, *Skandinavisk Aktuarietidskrift*, 1924, pp. 153—174 et *Comptes Rendus*, **179** (1924), pp. 1031—1034.

E. FELDHEIM, Simmons tételének új bizonyítása és általánosítása, *Matematikai és Fizikai Lapok*, **45** (1930), pp. 99—113; Nuova Dimostrazione e generalizzazione di un teorema del calcolo delle probabilità, *Giornale dell'Istituto Italiano degli Attuari*, **10** (1939), pp. 229—243.

Finalement, à l'aide de considérations de probabilités nous avons déduit les formules (9) et (14) utiles pour les fonctions Béta-incomplètes et les formules (17) et (18) utiles pour les fonctions Γ-incomplètes.

Traduisons ce que nous venons de dire en formules. La probabilité de l'événement favorable étant p à chaque épreuve, la probabilité d'obtenir ν cas favorables en n épreuves est

$$(1) \qquad P(\nu) = \binom{n}{\nu} p^\nu q^{n-\nu} \qquad \text{où} \qquad q = 1 - p.$$

La moyenne des cas favorables est np; l'écart observé est $\xi = \nu - np$. Le maximum de $P(\nu)$ a lieu pour ν égal au plus grand entier contenu dans $np + p$. D'après le théorème de SIMMONS, lorsque np est entier et $p < \frac{1}{2}$, on a

$$\sum_{\nu=0}^{np-1} P(\nu) > \sum_{\nu=np+1}^{n} P(\nu).$$

[Si $p > \frac{1}{2}$, c'est l'écart positif qui est plus probable.]

Nous examinerons le cas où np est quelconque, et nous poserons $np = k + \varepsilon$, où $k = [np]$ et $0 \leqq \varepsilon < 1$. Nos théorèmes se formulent ainsi:

I. Lorsque ε est plus grand ou égal à q, la probabilité de *l'écart positif est plus grande* que celle de l'écart négatif. Remarquons que, dans ce cas, le maximum de $P(\nu)$ se trouve du côté des écarts positifs.

II. Lorsque ε est inférieur ou égal à un demi, la probabilité de *l'écart négatif est plus grande* que celle de l'écart positif. Le maximum de $P(\nu)$ est situé du côté des écarts négatifs.

III. Lorsque ε se trouve entre $\frac{1}{2}$ et q, alors, jusqu'à une certaine valeur de ε, l'écart négatif est plus probable que l'écart positif; au delà de cette valeur, l'écart positif devient plus probable que l'écart négatif. Le maximum de $P(\nu)$ se trouve dans les deux cas du côté des écarts négatifs. Nous n'avons réussi à déterminer cette valeur limite de ε que d'une manière approchée; elle est approximativement $\frac{1}{2}\left(q + \frac{1}{2}\right)$.

I. Cas de $\varepsilon \geqq q$.

De la formule (1) on tire

$$P(\nu+1) - P(\nu) \equiv \Delta P(\nu) = \frac{np - \nu - q}{q(\nu+1)} P(\nu)$$

ce qui donne pour $\nu = k$

$$\Delta P(k) = \frac{\varepsilon - q}{q(k+1)} P(k).$$

Lorsque $\varepsilon = q$, la quantité $np + p = k + \varepsilon + p = k + 1$ est un entier et le maximum de $P(\nu)$ a lieu pour $\nu=k+1$; de plus, dans ce cas, la formule précédante donne $\Delta P(k)=0$, c'est-à-dire

$$P(k)=P(k+1). \tag{2}$$

Lorsque $\varepsilon>q$, le plus grand entier contenu dans $np+p$ est encore $k+1$, le maximum de $P(\nu)$ est $P(k+1)$, il est situé du côté des écarts positifs et l'on a

$$P(k)<P(k+1). \tag{3}$$

De la formule (1) on peut déduire

$$P(k+2+i)=\frac{(n-k-1-i)\,p}{(k+2+i)\,q}\,P(k+1+i), \tag{4}$$

$$P(k-1-i)=\frac{(k-i)\,q}{(n-k+i+1)\,p}\,P(k-i). \tag{5}$$

Examinons si le facteur dans la formule (4) est plus grand que celui qui figure dans (5). Ceci a lieu lorsqu'on a

$$[(n-k)^2-(i+1)^2]\,p^2>[(k+1)^2-(i+1)^2]\,q^2$$

ou (en éliminant k à l'aide de $k=np-\varepsilon$)

$$\left(n+\frac{\varepsilon}{q}\right)^2+(i+1)^2\left(\frac{1}{p^2}-\frac{1}{q^2}\right)>\left(n+\frac{1-\varepsilon}{p}\right)^2$$

et, à plus forte raison, lorsque

$$\frac{\varepsilon}{q}\geqq\frac{1-\varepsilon}{p} \qquad \text{ou} \qquad \varepsilon\geqq q.$$

Par suite, dans le cas considéré, le premier facteur est plus grand que le second, pour $i=0, 1, 2, \ldots$ En partant de (2) ou de (3) on déduit pas à pas que

$$P(k+1+i)>P(k-i),$$

donc

$$\sum_{i=0}^{k} P(k+1+i)>\sum_{i=0}^{k} P(k-i). \tag{6}$$

Pour démontrer que $n-1-k>k$, éliminons k ; on trouve

$$n+1>2(np-\varepsilon). \tag{7}$$

Montrons d'abord que

$$n-1>2(np-q)\,; \tag{8}$$

en simplifiant cela donne $nq+q>np+p$, ou $q>p$. On en conclut que l'inégalité (8) est satisfaite ; il en résulte

$$n+1>2(np+p)$$

et, par suite l'inégalité (7) est satisfaite à plus forte raison. Donc, d'après (6).

$$\sum_{i=0}^{n-k-1} P(k+1+i)>\sum_{i=0}^{k} P(k-i).$$

De là, le Théorème I: Lorsque $\varepsilon \geqq q$, la probabilité des écarts positifs est plus grande que celle des écarts négatifs. Le maximum de $P(\nu)$ se trouve du côté des écarts positifs (si $\varepsilon > q$) ou des deux côtés (si $\varepsilon = q$).

Remarque 1. La probabilité pour que le nombre des cas favorables ne dépasse pas λ en n épreuves est donnée aussi, d'une manière rigoureuse, par le rapport d'une fonction Béta-incomplète à la fonction complète correspondante:[2])

$$\sum_{\nu=0}^{\lambda} P(\nu) = I_q(n-\lambda;\ \lambda+1).$$

En posant $\lambda = k$, on obtient la probabilité d'un écart négatif.

D'après le théorème précédent, on a

$$I_q(n-k;\ k+1) = I_q(nq+\varepsilon;\ np+1-\varepsilon) < \frac{1}{2} \quad \text{si} \quad 1 > \varepsilon \geqq q > \frac{1}{2}.$$

Cette probabilité diminue quand ε augmente; la valeur la plus rapprochée de $\frac{1}{2}$ sera obtenue pour $\varepsilon = q$; alors on a

$$I_q[(n+1)q;\ (n+1)p] < \frac{1}{2}$$

pour toutes les valeurs de n, si $q > \frac{1}{2}$.

Cette inégalité sera avec la notation des tables de PEARSON

(9) $$I_x(\mathfrak{p}, \mathfrak{q}) < \frac{1}{2} \quad \text{si} \quad \mathfrak{p} \leqq \frac{1-x}{x}\mathfrak{q} \quad \text{et} \quad x \geqq \frac{1}{2}.$$

Ce résultat obtenu à l'aide des considérations de probabilités, est intéressant au point de vue des fonctions Béta-incomplètes.

II. Cas où $0 \leqq \varepsilon \leqq \frac{1}{2}$.

Dans ce cas, le maximum correspond à $\nu = k$; par suite

$$P(k) > P(k+1).$$

Au contraire, l'inégalité

$$P(k-i) > P(k+1+i)$$

n'est pas toujours vérifiée. Pour les petites valeurs de i, le facteur dans (4) est plus petit que celui qui se trouve dans (5), mais si i est assez grand le premier devient plus grand que le second, c'est-à-dire, pour une certaine valeur $i = \omega$,

$$P(k-\omega) < P(k+1+\omega).$$

---

[2]) CH. JORDAN, Calculus of Finite Differences (Budapest, 1939), p. 86.
K. PEARSON, Tables of the Incomplete Beta-Function (London, 1934).

De ce qui précède, il suit qu'en tout cas $\omega>0$. Si cette inégalité est satisfaite, on déduit de (4) et de (5) que, pour toutes les valeurs de $i>\omega$, on a

$$P(k-i)<P(k+1+i).$$

Dans ce qui suit nous utiliserons une méthode semblable à celle qui a été employée par M. FELDHEIM pour démontrer le théorème de SIMMONS dans le cas où np est entier.

Lorsque $i<\omega$, on a

$$(\omega-i)\,P(k+1+i)<(\omega-i)\,P(k-i)$$

et lorsque $i\geqq\omega$

$$(i-\omega)\,P(k+1+i)\geqq(i-\omega)\,P(k-i)\,;$$

on en tire

$$(10)\qquad \sum_{i=0}^{\omega-1}(\omega-i)\,P(k+1+i)<\sum_{i=0}^{\omega-1}(\omega-i)\,P(k-i),$$

$$\sum_{i=\omega}^{k}(i-\omega)\,P(k+1+i)>\sum_{i=\omega}^{k}(i-\omega)\,P(k-i).$$

Comme $n-k-1>k$, on a à fortiori

$$(11)\qquad \sum_{i=\omega}^{n-k-1}(i-\omega)\,P(k+1+i)>\sum_{i=\omega}^{k}(i-\omega)\,P(k-i).$$

En retranchant membre à membre l'inégalité (11) de l'inégalité (10), on obtient

$$(12)\qquad \sum_{i=0}^{n-k-1}(\omega-i)\,P(k+1+i)<\sum_{i=0}^{k}(\omega-i)\,P(k-i).$$

La moyenne des écarts étant nulle on a

$$\sum_{\nu=k+1}^{n}(\nu-np)\,P(\nu)=\sum_{\nu=0}^{k}(np-\nu)\,P(\nu).$$

En posant dans le premier membre $\nu=k+1+i$ et dans le second $\nu=k-i$, on trouve

$$(13)\qquad \sum_{i=0}^{n-k-1}(i+1-\varepsilon)\,P(k+1+i)=\sum_{i=0}^{k}(i+\varepsilon)\,P(k-i).$$

Finalement, en ajoutant (12) et (13), on a

$$\sum_{i=0}^{n-k-1}(\omega+1-\varepsilon)\,P(k+1+i)<\sum_{i=0}^{k}(\omega+\varepsilon)\,P(k-i),$$

et on en conclut que si $\varepsilon\leqq\frac{1}{2}$, on a

$$\sum_{i=0}^{n-k-1}P(k+1+i)<\sum_{i=0}^{k}P(k-i).$$

Théorème 2. La probabilité de *l'écart négatif est plus grande* que celle de l'écart positif, si $0 \leqq \varepsilon \leqq \frac{1}{2}$ en supposant $p < \frac{1}{2}$; par contre si $p > \frac{1}{2}$, l'écart positif est plus probable que l'écart négatif.

Remarque 2. En exprimant la probabilité de l'écart négatif par une fonction Béta-incomplète, on trouve $I_q(nq+\varepsilon; np+1-\varepsilon) > \frac{1}{2}$. Dans ce cas, la valeur la plus rapprochée de $\frac{1}{2}$ est donnée par $\varepsilon = \frac{1}{2}$; ce qui donne, avec les notations des tables de PEARSON, $\left(x > \frac{1}{2}\right)$:

$$I_x(\mathfrak{p}, \mathfrak{q}) > \frac{1}{2} \quad \text{si} \quad \mathfrak{q} \geqq \frac{1-x}{x}\mathfrak{p} + \frac{x-\frac{1}{2}}{x}. \tag{14}$$

Pour obtenir une valeur de $\mathfrak{q}$ donnant d'une manière approchée $I_x(\mathfrak{p}, \mathfrak{q}) = \frac{1}{2}$, on peut prendre la moyenne des grandeurs $\mathfrak{q}$ figurant dans les inégalités (9) et (14). On aura

$$\mathfrak{q} = \frac{1-x}{x}\mathfrak{p} + \frac{x-\frac{1}{2}}{2x} \tag{15}$$

ce qui correspond dans le problème de probabilité à $\varepsilon = \frac{1}{2}\left(q + \frac{1}{2}\right)$.

D'après les tables, généralement le premier entier supérieur à la valeur (15) de $\mathfrak{q}$ donne: $I_x(\mathfrak{p}, \mathfrak{q}) < \frac{1}{2}$, et le premier inférieur $I_x(\mathfrak{p}, \mathfrak{q}) > \frac{1}{2}$.

Exemple. $I_{0,8}(50; 12) = 0{,}4236$, $I_{0,8}(50; 13) = 0{,}5254$. Dans ce cas, $\mathfrak{q}$ donné par (15) est 12,518.

III. Cas où $\frac{1}{2} < \varepsilon < q$.

Lorsque ε varie de $\frac{1}{2}$ à q, alors jusqu'à une certaine valeur la probabilité de l'écart négatif reste plus grande qu'un demi et, au delà de cette valeur, elle devient plus petite. Dans les deux cas, le maximum de $P(\nu)$ correspond à $\nu = k$; elle est donc située du côté des écarts négatifs.

La valeur limite en question est d'après la *Remarque 2* approximativement égale à $\varepsilon = \frac{1}{2}\left(q + \frac{1}{2}\right)$. On en conclut que si $\varepsilon \geqq \frac{q+\frac{1}{2}}{2}$, la probabilité de l'écart négatif est généralement inférieure et dans le cas contraire supérieure à un demi.

Les Tables de la fonction Béta-incomplète montrent que si $q < 0{,}9$ la valeur ci-dessus donne des résultats exacts, mais si $q \geqq 0{,}9$ il faut retrancher de $\frac{1}{2}\left(q+\frac{1}{2}\right)$ la fraction 0,045. Par suite, lorsque $q \geqq 0{,}9$, la probabilité de l'écart négatif sera plus grande que la probabilité de l'écart positif si

$$\varepsilon < \frac{1}{2}\left(q+\frac{1}{2}\right) - 0{,}045. \tag{15'}$$

Exemple 1. $p=0{,}1$, $q=0{,}9$, $\frac{1}{2}\left(q+\frac{1}{2}\right)=0{,}7$

n	k	ε	$I_q(n-k, k+1)$
6	0	0,6	$I_{0,9}\ (6;1) = 0{,}5314$
7	0	0,7	$I_{0,9}\ (7;1) = 0{,}4783$
16	1	0,6	$I_{0,9}(15;2) = 0{,}5147$
17	1	0,7	$I_{0,9}(16;2) = 0{,}4818$
26	2	0,6	$I_{0,9}(24;3) = 0{,}5105$
27	2	0,7	$I_{0,9}(25;3) = 0{,}4846$
36	3	0,6	$I_{0,9}(33;4) = 0{,}5085$
37	3	0,7	$I_{0,9}(34;4) = 0{,}4864$
46	4	0,6	$I_{0,9}(42;5) = 0{,}5073$
47	4	0,7	$I_{0,9}(43;5) = 0{,}4878$

On voit que les probabilités s'approchent d'un demi si n augmente.

Exemple 2. $p=1/5$, $q=4/5$, $\frac{1}{2}\left(q+\frac{1}{2}\right)= 13/20$.

33	6	12/20	$I_{0,8}(27;7) = 0{,}5004$
34	6	16/20	$I_{0,8}(28;7) = 0{,}4661$

Exemple 3. $p=0{,}3$, $q=0{,}7$ $\frac{1}{2}\left(q+\frac{1}{2}\right)=0{,}6$

35	10	0,5	$I_{0,7}(25;11) = 0{,}5100$
32	9	0,6	$I_{0,7}(23;10) = 0{,}4951$

Exemple 4. $p=1/20$, $q=19/20$, $\frac{1}{2}\left(q+\frac{1}{2}\right)= 29/40$

n	k	ε	
33	1	26/40	$I_{0,95}(32;2) = 0{,}5036$
34	1	28/40	$I_{0,95}(33;2) = 0{,}4872$
35	1	30/40	$I_{0,95}(34;2) = 0{,}4567$

Exemple 5. $p=0{,}02$, $q=0{,}98$ et $\frac{1}{2}\left(q+\frac{1}{2}\right)=0{,}74$

34	0	0,68	$I_{0,98}(34;\,1)=0{,}5031$
35	0	0,70	$I_{0,98}(35;\,1)=0{,}4931$
36	0	0,72	$I_{0,98}(36;\,1)=0{,}4832$
37	0	0,74	$I_{0,98}(37;\,1)=0{,}4735$

Ces résultats sont conformes à la règle (15').

Remarque 3. Comme dans les Tables, $I_x(\mathfrak{p}, \mathfrak{q})$ n'est calculé que pour $\mathfrak{p}\leqq 50$ et $\mathfrak{q}\leqq 50$, et $\mathfrak{p}=n-k$, les formules précédentes sont difficilement utilisables lorsque $n>50$; mais si p est petit, on peut employer la formule approchée connue, donnant la probabilité d'un écart négatif à l'aide des fonctions Γ incomplètes[3])

$$(16)\qquad \sum_{\nu=0}^{k} P(\nu) \sim \sum_{\nu=0}^{k} \frac{(np)^{\nu} e^{-np}}{\nu\,!} = 1 - I(\bar{u},\, \mathfrak{p})$$

où $\bar{u}=\dfrac{np}{\sqrt{k+1}}=\sqrt{k+1}-\dfrac{1-\varepsilon}{\sqrt{k+1}}$ et $\mathfrak{p}=k$.

Lorsque $\varepsilon\leqq\frac{1}{2}$, la probabilité (16) et plus grande que $\frac{1}{2}$ et $I(u,\, \mathfrak{p})<\frac{1}{2}$. De la formule de DOODSON[3]) on conclut que

$$(17)\qquad I(\bar{u},\, \mathfrak{p})<\frac{1}{2} \quad \text{si} \quad \bar{u}\leqq\sqrt{\mathfrak{p}+1}-\frac{\frac{1}{3}}{\sqrt{\mathfrak{p}+1}}.$$

(Cette valeur est plus serrée que celle qui se trouve dans loc. cit. 2.)

Lorsque $\varepsilon\geqq q$, la probabilité (16) est plus petite que $\frac{1}{2}$ et $I(\bar{u},\, \mathfrak{p})>\frac{1}{2}$. Il en résulte à fortiori, en négligeant le terme $1/3\sqrt{\mathfrak{p}+1}$,

$$(18)\qquad I(\bar{u},\, \mathfrak{p})>\frac{1}{2} \quad \text{si} \quad \bar{u}\geqq\sqrt{\mathfrak{p}+1}.$$

Les tables de la fonction Γ incomplète sont calculées jusqu'à $\mathfrak{p}\leqq 50$. Comme $\mathfrak{p}=k=np-\varepsilon$, et que p est petit, n peut être déjà assez grand.

Exemple 1. Soit $p=0{,}1$, $q=0{,}9$ et $n=47$. *Formule rigoureuse*. $P=I_q(\mathfrak{p}, \mathfrak{q})$. On a $np=4{,}7$, $k=4$, $\varepsilon=0{,}7$, $\mathfrak{p}=44$, $\mathfrak{q}=5$. Comme

$$\varepsilon>\frac{1}{2}\left(q+\frac{1}{2}\right)-0{,}045=0{,}655,$$

la probabilité de l'écart négatif doit être d'après (15') plus petite qu'un

3) K. PEARSON, Tables of the Incomplete Γ-Function (London, 1922), p. VIII; A. T. DOODSON, Biometrika, Vol. XI. p. 428.

demi. On trouve

$$P = I_{0,9}(44,5) = 0,48778.$$

Formule approchée. $P = 1 - I(\bar{u}, \mathfrak{p})$. On a $\bar{u} = np/\sqrt{k+1} = 2,1466$. $\mathfrak{p} = k = 4$. Comme dans ce cas

$$\bar{u} > \frac{3(k+1)-1}{3\sqrt{k+1}} = 2,087,$$

on doit avoir, d'après (17), $I(\bar{u}, \mathfrak{p}) \geqq \frac{1}{2}$. On trouve $I(u, p) = 0,50468$. où $P = 0,49532$.

Exemple 2. Soit $p = 0,1$ et $n = 46$. *Formule rigoureuse.* On a $np = 4,6$, $k = 4$, $\varepsilon = 0,6$, $\mathfrak{p} = 42$, $\mathfrak{q} = 5$. Comme $\varepsilon < 0,655$, la probabilité de l'écart négatif doit être, d'après (15'), plus grande qu'un demi; en effet on a

$$P = I_{0,9}(42,5) = 0,50732.$$

Formule approchée. On a $np = 4,6$, $\mathfrak{p} = k = 4$ $\bar{u} = np/\sqrt{k+1} = 2,0572$. Comme $\bar{u} < 2,087$ on doit avoir, d'après (17), $I(\bar{u}, \mathfrak{p}) < \frac{1}{2}$. On trouve $I(2,057; 4) = 0,48663$ et $P = 0,5134$.

(Reçu le 4 novembre 1942.)